知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

科学数据开放共享中出版商政策研究

◎ 崔雁[1,2]

[1] 北京交通大学图书馆 北京 100044
[2] 中国科学院文献情报中心 北京 100190

摘要：[目的 / 意义] 在科学数据的开放共享体系中，出版商是关键利益相关者。出版商与包括资助者、数据中心与知识库及研究机构在内的其他利益相关者合作，共同促进形成开放共享的研究数据文化。尽管出版商已经意识到科学数据开放共享的重要价值，但出版商参与开放共享研究数据的实践，目前还未形成规模。[方法 / 过程] 梳理目前出版商有关开放共享数据的政策，分析其共性与差距，对目前科学数据共享实践中，出版商的政策进行分析。[结果 / 结论] 针对出版商提出相关政策建议，有助于出版商进一步推动实现开放共享研究数据的目标。

关键词：科学数据开放共享 出版商 政策
分类号：G251

引用格式：崔雁. 科学数据开放共享中出版商政策研究 [J/OL]. 知识管理论坛 , 2016, 1(3): 182-191[引用日期]. http://www.kmf.ac.cn/paperView?id=35.

科学数据的开放共享，具有重要意义，有利于经济价值与学术价值的提升。正如欧盟委员会所指出，“科学数据的开放获取，提高了数据质量，减少了重复研究的成本，加快了科技进步，并有助于打击学术造假”[1]。英国皇家学会报告认为，开放对于科学事业具有重要价值，开放可以识别错误，支持、反对或优化理论，数据的重用可更深入了解和掌握知识 [2]。开放数据是开放科学的一部分，具有可获得、可理解、可评估和可用的特征，开放数据与科学出版物的开放获取相结合，可以有效沟通与交流科学的研究内容。

1 科学数据开放共享中的出版商

1.1 科学数据开放共享

在科学研究过程中，有不同的利益相关者参与其中。这些利益相关者类别多样，性质各异，共同形成科学研究的生态系统。根据各利益相关者的职能，围绕研究人员，将这一系统进行分类研究，确定出 4 个利益相关者群体，即资助者、数据管理者、研究机构与出版商作为关键参与者，以促进和实现科学研究数据的开放共享。

在这一系统中，每类利益相关者都有各自不同的职能、驱动力及利益点，但总体利益与动机相同，即促进科学的进步。利益相关者的利益与行为动机与各自职能连接在一起，形成了科学研究系统的边界。该系统本质上受到外部和内部因素的双重影响：外部因素可能表现为政治、社会、经济、科学和文化体系等；内部因素包括政治意愿、经济及学术竞争、技术基础设施、法律、道德等因素。

作者简介：崔雁 (ORCID：0000-0002-2577-573X)，馆员，博士研究生，E-mail：ycui@bjtu.edu.cn。
收稿日期：2016-04-21 发表日期：2016-05-27 本文责任编辑：王传清

知识管理论坛，2016(3):182–191
DOI：10.13266/j.issn.2095–5472.2016.023

该系统的价值与动机是外部和内部因素共同作用的结果，其中外部因素控制资源输入到该系统中，内部因素控制这些资源的可用性和分配。如图 1 所示：

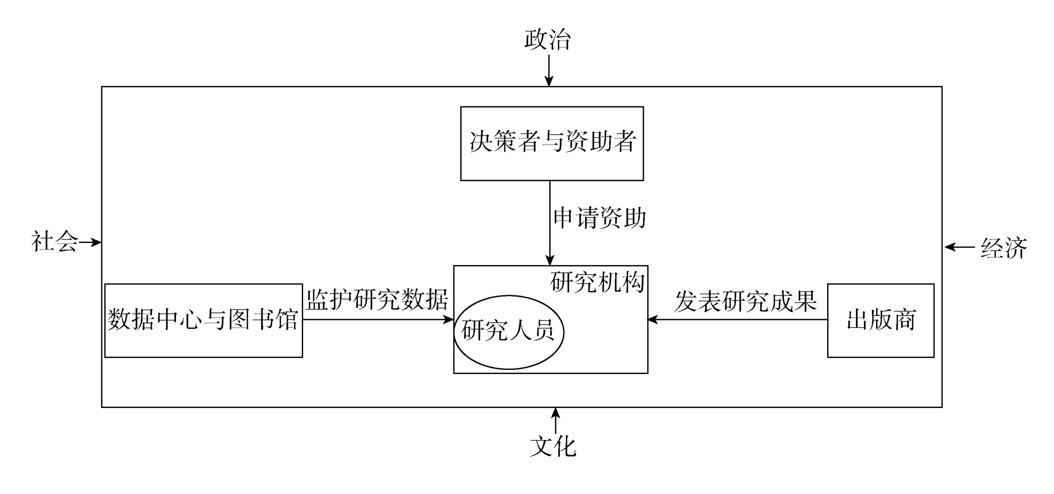


图 1 科学研究系统示意

在此科学开放系统中，利益相关者群体职能众多并偶有重叠，总体上各利益相关方对研究数据的开放获取的意义已达成共识，但各利益相关方对实现开放研究数据的方式的认识并不一致。本研究主要探讨在推动科学数据开放共享中，出版商数据政策的现状、问题，并构建相应模型，提出相应建议。

1.2 出版商在科学数据开放共享系统中的目标

在促进研究数据开放共享方面，各利益相关方发挥不同的作用。出版商通过期刊向作者提供数据共享政策，在数据开放中发挥着重要作用。研究出版商的数据政策，是因为这是在研究过程中将数据政策与研究人员密切关联起来的点，为发布研究成果，研究人员有动力、有可能遵守数据政策。同时，研究发现，虽然包括出资者和研究机构都可能有适用于研究人员的政策，但遵守这些政策的研究人员的比例还很低。研究人员没有遵守数据政策的直接动力；此外研究人员在需要进行数据存档时，可能也没有合适的机构知识库可以选择。出版商提出的数据政策，是在研究人员完成研究过程后发表研究成果前，研究人员有可能、有动力遵守出版商的数据政策。

1.2.1 科学数据开放共享的价值

科学的思想与科学的证据互相佐证，研究数据和科学实践之间的联系是不言自明的，因此，推动开放获取研究数据，与促进科学严谨的目标相一致。更大程度地共享与获取数据，能加强科学的基础价值，即允许科学研究有效地自我校正，以便二次分析、检验、质疑或改进原有成果 [3]。此外，开放获取政策通过减少重复工作，最大限度地减少研究人员用在搜索信息和数据方面的精力，因而改善了他们的工作条件。

JISC 以英国高等教育和研究机构为例，指出研究数据的监护与开放共享具有如下利益 [4]：①研究人员能够更广泛地获得数据，从而促进更多的跨部门的合作，研究人员在行业、政府和非政府组织内有可能获得大量的教育和培训机会。数据的开放共享有助于实现数据使用和重用，降低数据的收集和复制成本，分担数据收集的直接和间接成本（如避免调查疲劳，从而提高响应率等），创造出在数据收集时未曾预想到的新的利用方式，进行数据挖掘等。②在项目申请拨款和评估阶段、出版和研究评估阶段，更容易发现欺诈和抄袭，更容易进行评估和同行审

查。因此将有机会创造更加完整和透明的科学记录。③通过将研究人员、知识库、资助者与有价值的资源相链接，从而有更多的机会提高研究的可见度。

科学研究系统中的各个利益相关者群体认识到走向开放数据的利益，认同将数据的开放共享置于战略高度进行考虑。科学表现为知识积累的过程，数据在促进早期工作中起着重要作用。开放获取研究数据，通过避免重复劳动，促进协作，有助于显著加快这一科学过程，从整体上使科学成为更加透明的进程，实现推动公众参与，激励创新和改革公共服务的宗旨[5]。

1.2.2 基本情况

出版商作为利益相关者，越来越关注数据密集型研究，认同开放获取研究数据的意义。很多出版商认同布鲁塞尔宣言，即"所有的研究者应能自由地获取原始研究数据。出版商鼓励公开公布原始研究数据结果，将相关的数据集及子数据集与论文共同提交给期刊，应尽可能让其他研究人员自由获取"[6]。很多出版商的数据政策为强制性政策，要求作者将支持出版物的研究数据存储在经认证的知识库中，实现开放获取。研究表明，如果期刊具有强制性数据政策，并且有可获取数据的声明，那么在线找到该数据的可能性，几乎是没有类似政策的期刊的 1 000 倍[7]。

尽管目前出版商已开始重视出版物的开放获取，将开放获取作为一种出版的商业模式，不过出版商参与研究数据的出版，特别是开放获取研究数据，目前尚未形成规模。出版商关注研究数据及其开放，重要原因在于数据为出版商的主要产品即出版物增加了价值，数据有助于验证研究成果，从而增强了所发表研究成果的可信性，而可信度对于研究具有重要意义[2]。此外，资助机构的政策要求公开获取研究数据，实质上向出版商施加了压力，出版商需要参与到研究数据开放获取中，帮助作者和研究机构符合出资人的要求。现在，领先发展的出版商开始与其他利益相关方合作，试图挖掘研究数据的潜能，形成以数据为基础的新产品和服务，对研究数据进行同行评议，开展提高数据质量的其他服务。

2 出版商主要数据管理政策

本研究根据 STM 的年度报告[8]，确定出以下出版商为主要分析对象（见表 1）。这十大出版商所出版期刊占到 2014 年所有出版期刊的 45.2%，另外本研究将开放获取出版商，科学公共图书馆（PLoS）、生物医学中心（BioMed Central）的数据开放政策考虑在内。

表 1 以期刊数量而论的前 10 位出版商

序号	出版商	出版期刊数量（种）	占所有期刊的累积百分比（%）
1	Springer	2 987	10.6
2	Elsevier	2 500	19.5
3	Wiley	2 388	28
4	Taylor & Francis	2 105	35.5
5	SAGE	750	38.1
6	Wolters Kluwer	672	40.5
7	Hindawi	438	42.1
8	CUP	350	43.3
9	OUP	362	44.6
10	NPG(包括 Frontiers)	178	45.2

知识管理论坛，2016(3):182–191
DOI：10.13266/j.issn.2095–5472.2016.023

2.1 出版商数据政策现状

2.1.1 将开放获取作为默认情况

出版商将数据的开放获取作为订立政策的基础，如 PLOS 研究数据开放获取强制性政策 [9]，指出除了极少数例外情况，支持 PLOS 出版物的所有研究数据都必须开放获取。作者在向 PLOS 提交稿件时，要同时提交数据可用性声明，在声明中表明遵守 PLOS 的政策规定，在手稿成功提交后，数据作为最终手稿的部分内容发布。PLOS 要求作者将数据存储于推荐的经认证的数据中心或知识库。小数据集可与稿件一并上传。PLOS 的编辑和投稿指南，向研究人员提供指导，协助研究人员遵守期刊开放数据政策。在限制数据获取的情况下，PLOS 有权发布修正说明，联系作者的机构或资助者，甚至撤销出版。

2.1.2 新的出版形式的出现

出版界越来越关注开放数据，产生了一种新型出版产品，即数据期刊。数据期刊的出现，与数据可以单独发布紧密相关。数据的单独发布可以确保数据作为科学记录的基本组成部分，以可理解的形式向科学界提供。数据期刊是同行评议的开放获取平台，用于发布、分享和传播各学科的数据。发表的数据论文包含数据集的具体相关信息，如收集、处理方式等。发表的数据论文与认可的知识库互相关联，数据论文引用存储于知识库或数据中心中的数据集。正如澳大利亚国家数据服务中心（ANDS）在其数据期刊指南中所指出的，“从根本上说，数据期刊寻求促进科学认证和再利用，提高科学方法和结果的透明度，支持良好的数据管理方法，并为数据集提供一个可访问的、永久的、可解析的路径”。ANDS 指出，数据论文的出版过程包括对数据集的同行评议，最大限度地提高了数据再利用的机会，并为研究人员提供了学术认可的可能性 [10]。

2.1.3 同行评审

对研究数据与数据出版物开放获取的关注，彰显了研究数据的科学质量及研究数据同行评审的重要意义。在此背景下，一些出版商将同行评议的范围扩大到包括数据在内的同行评审。M. S. Mayernik 等 2014 年进行了有关“数据同行评审”的研究，提出因为出版物或资源类型有所不同，进行同行评审的方式也必须有所变化。研究者对几种类型的评审资源进行了区分，包括在传统科学论文中分析的数据，在传统科学期刊上发布的数据文章，以及通过数据期刊发布的开放获取知识库与数据集。M. S. Mayernik 等针对数据的同行评审、数据的质量保证过程确定了一些共同因素，包括：可通过数据中心或知识库获取数据集；数据集有足够的信息以备评审；期刊有明确的方针指明审核的要点，指导评审者进行数据审查等 [11]。

开放考古学杂志（JOAD[12-13]）对所有提交的数据论文采用同行评审程序，评审内容包括论文的内容与存储的数据。论文的内容指与数据集的建立和重用相关的信息，以及对数据集的描述。存储的数据指以可持续性模式提交到存储库的数据，包括其许可方式。

2.1.4 数据引用

除对数据进行同行评审外，出版商还逐步引入数据引用政策，以促进研究数据的标准化使用。研究数据对研究过程具有重要价值与意义，为扩大高质量研究数据的传播，形成数据利用的规范方式，FORCE11[14] 制定了数据引用的主要原则。FORCE11 的引用原则的前提是数据引用需要实现人类和机器均可读。该数据引用原则可能并不全面，主要目的是鼓励各学科制定体现自身特点的引用方式。

FORCE11 原则包括：说明数据引用的重要性；通过数据引用促进学术信用；数据引用要实现机器可操作性，包括全球永久唯一标识符；数据引用要促进对数据本身的获取；数据引用应该具有持久性；便于识别，易于获取，可以验证；具有互操作、灵活性等特点。



2.1.5 内容发现和链接服务

出版商逐渐把发展方向转移到内容发现和链接服务，专注于文本与数据挖掘（TDM）工具，以便能开发内容，进而提供相关服务。出版商越来越关注数据挖掘是研究人员利用大型数据库的内容、数据和出版物的需求的直接结果。一份文本和数据挖掘专家小组的报告指出，TDM 是一种重要技术，可用于从指数级增长的数字数据中，分析和提取新的见解和知识[15]。该报告的结论是，因为研究人员的技能和技术不断提升，所研究的数据集的复杂性、多样性及规模不断扩大，因此 TDM 有可能会更加重要。但对于利用文本与数据挖掘工具是否应有所限制，目前仍有争议。

2.2 期刊数据政策问题

2.2.1 数据格式与文件大小的差异

期刊要包括支持文章结果的所有数据，往往是很难实现的。研究的方法不同，产生的数据也大不相同，数据的格式和文件大小差异巨大。定性研究生成的数据，多以文本形式存在，例如实地观察笔记，或采访或报道的文字记录等。定量研究生成的数据，多以电子表格的形式保存。一项研究可能产生多种类型的数据，而论文可能包括附加文本、数值数据集和数字图像，这些都可能增加论文的大小。因此出版商表现出对集成到每篇论文中的数据集大小的关注。某些出版商开始尝试出版在线期刊文章，以包括多种数据，例如爱思唯尔的有关未来的文章的探索[16]。然而，并不是每一种期刊都有包括各种数据的能力。这就要求期刊的办刊方针应清楚说明，作为论文组成部分的数据，在何种程度上可以包括在论文中。

2.2.2 机构知识库的成本

为解决出版商服务器超载的问题，将期刊文章的重要数据链接到一个特定的机构知识库，可能是一个合理的选择，但这将相关的长期运营成本转嫁到了机构中。但资助者目前的基金中并不包括这部分资金，而机构可能也并不愿意在当前的管理费用中增加这种支出。这就使得在机构知识库中存储科学数据的可持续性有待探讨。

2.2.3 研究人员对数据开放的认识

许多研究人员并不知道相关的知识库，为此期刊数据政策应说明，数据是否应该在认可的知识库中存储，是否要使用永久统一资源定位符（URL），是否要采用某种形式的数据引用。数据发布的时间也是一个需要关注的问题，研究人员并不关心在出版过程中的什么时候数据可被公开访问，而是关心在研究过程中何时数据应被公开访问。研究论文并不是在研究结果全部产生后才会形成，而是在研究过程中逐步产生的。在论文发表的同时，是否适合发布研究数据取决于多种因素，诸如某些形式的数据有敏感性，要保护受试者等因素。

3 构建期刊研究数据策略模型框架

3.1 出版商期刊政策的基本要求

有效的政策制定过程必然需要将注意力集中于数据共享过程中各利益相关方的意见，而目前的数字基础设施在不断变化，出版商、知识库和系统之间，并没有强有力的措施鼓励共享数据。共同点包括：①出版商共享数据的方式差异很大；②在出版过程中，出版商对所接受的数据类型、数据应存储的地点、应存储的时间等说明，模糊不清；③研究人员普遍赞成共享数据，但研究人员不知道该如何克服共享障碍；④研究人员认为出版商和期刊有关数据格式和存放地点等清晰的政策，将有益于研究；⑤出版商也认为在数据的关联与嵌入方面存在障碍。

科学数据共享的许多问题，能够在出版过程中通过期刊强有力而明确的政策加以解决。因此，本研究的目标是确定能推荐给期刊使用的政策模型。研究收集到的期刊政策信息，对材料进行分析后，归纳出目前主要的政策要求，如表 2 所示：

知识管理论坛，2016(3):182–191
DOI：10.13266/j.issn.2095–5472.2016.023

表 2 期刊数据政策的基本要求

序号	政策要求
1	政策应有一般性声明，列举数据共享的益处
2	政策应清楚地说明，属于期刊、出版商或者专业协会的政策
3	文章中包括或可链接的数据类型
4	数据的格式，覆盖所有学科领域准则
5	相关的数据说明，如数据引用，其他元数据等
6	数据是否要求或建议共享，以及对数据数量的限制
7	根据数据类型，确定数据存储的地点
8	凡注明数据是可用的，说明如何获取
9	在出版过程中应提供数据
10	是否有时滞期及时滞期的长度
11	数据是否可以公开获取、自由获取、低成本获取或有其他级别的限制
12	作者应声明数据可重用的所有情况与条件
13	是否有数据政策的例外情况
14	有对作者遵守政策的监管方式
15	对作者说明不遵守政策的后果
16	在其他研究人员投诉无法满足数据请求时，期刊相关的处理程序说明

3.2 框架模型

基于以上所列出的观察，形成基本的期刊研究数据政策的模型框架。由资助者和研究机构作为一方，与出版商的另一方进行合作，发展数据政策。表 3 为所构建的期刊研究数据策略模型框架。

表 3 期刊研究数据策略模型框架

政策	元素	内容说明	具体表述方式
总体要求	目的	说明数据开放共享的益处	本出版社认为，支持文章得出结论的数据应为所有研究团体可用，以促进相互合作，证明研究结果的可验证性，并鼓励复用与再利用数据。本出版社认为，通过对作者工作成果在更大范围的开放，扩大数据的透明度，将提高引用率与提高科学质量，最终价值将有利于作者
	范围	说明政策适用对象	①本研究数据政策适用于本学会的政策；②本研究数据政策适用于本期刊的编辑政策；③本研究数据政策适用于本出版社的政策
		明确说明，存储数据是出版的必要条件，还是仅建议存储数据	①所提交的论文得以出版的条件是，论文结论所基于的所有数据，必须由作者存储于指定位置，该位置可以自由公开获取数据，这是强制性要求；②建议将文章结论所基于的所有数据，由作者存储于指定位置，该位置可以自由公开获取数据；③与论文相关的数据不必公开获取
		在显著位置明确表示，希望作者遵守数据政策，并对包括论文出版与数据存储之间合理的时间区间作出规定	①本出版社希望所有作者遵守研究数据政策；②本出版社在收到存储库已正式存储数据的通知时，才会出版论文；③本出版社允许作者在数据公布一个月内进行数据的存储
		在因资助而产生研究的全部或部分数据时，政策应该要求作者说明研究的资助者	作者必须指明研究的资助者，说明文章中哪些数据集是基于资助产生的

（续表 3）

数据提交与评审	数据提交	明确地指出，数据是否可以作为论文的组成部分，以及论文中是否应包括超链接，或者说，在数据与论文保存在不同的地点时，是否通过超链接实现关联	①数据嵌入到发表的文章或附录中；②数据不必嵌入到发布的文章或附录中；③数据可通过文章中的超链接获取，这将引导至另一台服务器，该服务器可能由出版商所控制，也可能与出版商无关；④应对感兴趣的研究人员能够访问该数据作出相应安排
		考虑到必要数据和补充数据的区别，政策应该明确指明，所接受的数据的类型	①可链接或嵌入到文章中的数据，可接受的数据形式包括：视频图像、音频文件、软件、电子表格、基于文本的文件、DNA 序列等；②可链接或嵌入到文章中的数据，不可接受的数据形式包括：视频图像、音频文件、软件、电子表格、基于文本的文件、DNA 序列等
		清楚地表明接受的数据格式，并给出对数据保护的解释说明	①接受任何格式的数据；②只接受 ASCII 格式的数据，以实现数据的保存和互操作；③接受开放格式的数据，以实现数据的保存和互操作；④数据需要获取代码才能实现可复制，则数据与代码共同存储
	数据评审	明确数据是否应该进行评审，由何人进行评审	①数据应该与文章一起提交，以便由指定的评审小组进行同行评审；②数据应该独立进行审查；③不对数据进行审查
数据标准与指导	数据标准	提供元数据和作者标识符的信息	①数据集必须有一个整体的数字对象标识符（DOI）；②数据的每个项目都必须有 DOI；③数据应与相应描述的可阅读文件共同提交；④编码和软件、缩写和使用的术语，计量单位和任何其他相关联的数据的细节需与数据共同提交
	数据指导	对从较大的数据集中选择与发表的文章最相关的数据时，应当给予指导	①如果发表的文章是基于较大的数据集中选取的有限数据，仅需存储与论文密切必要的数据；②如果发表的文章是基于较大的数据集中选取的有限数据，要求整个数据集可以公开获取；③如果发表的文章是基于较大的数据集中选取的有限数据，作者可以选择存储部分或全部数据集
		对数据分析方法的公布给出指导	①在相关文章中应明确指出数据分析的方法；②应提供数据分析的详细方法，以实现研究的可复制性；③作者可以选择概述数据分析的方法
数据的保存与共享	数据保存	说明数据是否应存储在指定的存储库中，或由作者选择存储地点	①本期刊是综合性期刊，数据必须在指定的存储库存放，如 Dryad 等；②数据必须在由本协会认可的存储库中存放；③数据可以在具有认可资质的存储库中存放；④数据可以在主要作者所在的机构知识库中存放；⑤数据可存储于由作者自由决定的可信的存储库中；(6) 数据可以通过相应安排，方便研究人员获得数据
		数据如果与另一台服务器上的论文链接，数据政策应该明确应该使用的 URL 形式	①用于连接到数据的 URL 必须可永久链接；②用于连接到数据的 URL 必须是数字对象标识符；③作者可以使用统一资源标识符以链接数据到文章；④作者可以使用持久统一资源定位符以链接数据到文章
	数据引用	对数据引用作出说明，给出数据引用格式的指导	①不需引用数据；②作者可以选择引用数据；③数据应采用下面的方法进行引用，如 ……
	数据获取	说明公开发布数据是否有时滞期	①论文一发表，即可公开获取数据；②论文发表之前，可以公开获取数据；③论文发表至少某个阶段后，可以公开获取数据；④文章发表之前必须存储数据；⑤有时滞期的文章发表时，可以存储数据
		在政策允许某些类型的数据有例外获取的情况时，应明确说明例外的标准	期刊编辑委员会，认同研究数据有不适宜公开获取的情况，但作者需说明可能会造成的后果，如： ①研究参与者个人或家属的生命或生活会受到严重损害 ②对已成熟的行业造成严重后果 ③对国家安全造成严重后果等

（续表 3）

知识管理论坛，2016(3):182−191
DOI：10.13266/j.issn.2095−5472.2016.023

数据所有权与安全	数据所有权	要求作者说明相关数据或者数据的不同部分的知识产权状况	应明确说明再利用的范围等，例如，可使用广泛理解接受的 CC 许可，作为再使用规范进行说明。作者的声明应包括先前的知识产权及再利用的要求，适用于出资者或机构的政策，包括时滞期与敏感数据的处理问题，如：本数据是在资助者资助下产生的结果，作者、所属机构及资助者遵守相关知识产权政策。数据的发布符合资助政策，遵守“署名非商业相同方式共享（CC BY-NC-SA）”，6 个月时滞期后公布
	隐私保护	说明对人类受试者数据公布的伦理问题的处理方式	①在存储数据之前，人类受试者的数据标识如姓名、地址、出生日期、身份识别号码，电话号码等应当移除；②人类受试者及其他敏感数据，在数据发布之前允许有时滞期；③个别研究人员要获得人类主体和其他敏感数据，可由作者进行特定安排；④在特殊情况下，人类受试主体和其他敏感数据可被例外允许获取
		对作者提供指导，以使个体研究人员可以获取敏感数据	①在敏感数据不适于公开的情况下，作者应该作出安排，以便于个别研究人员获取数据；②在敏感数据不适于公开的情况下，作者应该作出安排，以便于个别研究人员可以复制研究；③在数据为敏感数据的情况下，可提供作者的联系方式给有关方面
数据监管	数据监管	明确列出不遵守期刊研究数据政策的后果及监管方式	①如果本期刊收到其他研究人员无法获取已出版论文相关数据的投诉，作者必须出示有关已存储数据的证明；②如果不遵守本协会的数据政策，本协会将撤销作者的会员资格；③如果相关数据未在规定时间期限内存储，本出版社将不会出版作者撰写的相关文章

3.3 实施方式

3.3.1 逐步制定出版物的支持性数据的强制性开放共享政策

出版商应该支持期刊编辑制定研究数据的强制性政策，从而提高研究过程透明度，扩展研究数据的潜力。数据的编辑政策应解决诸如文档、元数据、数据出版格式、许可、引用等问题。编辑政策要求作者在文章提交过程和同行评议过程中，提交数据的可用性描述。对提交的文章不符合规定的情况，政策应提出对应的措施，如撤销发表的文章。

3.3.2 与认证知识库和数据中心协作，简化数据提交流程

出版商可按照相关标准考查知识库与数据中心的可信度，与符合数据认证标准的数据中心与知识库协作。类似的知识库或数据中心可以是主题明确的学科知识库，也可以是机构知识库或综合知识库。二者的合作应该会产生积极影响，为研究人员和研究机构提供高品质的产品和服务，服务具有可操作性、技术上无缝的特点，从而实现以开放格式传播和保存高质量的学术产品和研究数据。对于没有公认的数据中心或知识库的学科，出版商对研究人员提供指导和帮助，提供适合存储与获取的机构知识库建议，或提供商业数据服务。在这方面，出版商的角色应该是更多地促使学术团体建立资源库评估准则，从而帮助研究人员选择合适的存储库。当学术界建立起知识库的认可标准，出版商就可以通过期刊政策执行这些标准。

3.3.3 数据作为一流学术成果，接受同行评议

出版商的主要职责之一是确保出版物的高质量，而研究数据同行评审制度的建立，有助于提升出版产品的品质。同行评审过程，应详细说明评审的标准，要评估研究数据的技术与质量问题，技术方面如考量数据集的完整性和一致性，收集数据的标准，使用的软件等；研究数据的科学质量则由研究团体通过出版前与出版后的同行评审进行评估。同行评审流程中，出版商也应该探索对评审者的激励方式，包括支付酬金，邀请编写特稿，加入编辑委员会，甚至聘用一些专家进行内容评审。

3.3.4 发展有关研究数据引用的策略

研究数据通过传统的出版过程传播，为实现数据的长期可重用，需建立并使用数据的引用标准。在这方面，出版商应要求出版物及相关的数据可引用，并为数据的引用提供明确指导。在说明数据引用时，出版商可参照一些已有的实践，如 Force11[14] 的原则，参与 DataCite[17]，加入到研究界和编辑讨论的过程中。数据引用应包括 DOI，以及使用许可信息，如创作共用许可，数据的引用方式最好是机器可操作的，可让用户知道可以如何处理研究数据。

3.3.5 建立许可政策鼓励进行文本数据挖掘

出版商的编辑政策应以清晰的方式，说明研究数据的发布与获取方式，如默认或建议开放获取等。考虑到文本数据挖掘工具可带来显著经济效益，鼓励出版商调整策略，允许研究人员在研究中使用这种技术。

4 结语

尽管出版商已经强烈关注出版物的开放获取，将开放存取作为一种商业模式，但出版商参与研究数据，特别是开放研究数据，还未形成规模。出版商关注研究数据与研究数据开放，是因为数据为出版商的主要产品即出版物增加了价值，通过对研究成果的验证功能，增强了所发表研究成果的可信性，而可信度对于研究具有重要意义。同时出版商为应对来自于出资人对开放获取的压力，也在逐步制定政策。

出版商和出版商的数据政策在更广泛地采用数据归档，与协助数据存档政策模型的发展过程中，可以发挥关键作用。出版商所提出的数据政策，应能体现透明性，实现重用的可能，提出存储数据的要求，提供对数据结构和元数据的指导，向作者指明适当的网络链接的存储库。数据政策不仅有利于研究人员自己及相同或相关领域同行的研究，同时促进了数据归档和数据链接，完善了数据的存储。

参考文献：

[1] European Commission. Recommendation on access to and preservation of scientific information [EB/OL]. [2016-03-20]. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[2] BOULTON G, CAMPBELL P, COLLINS B, et al. Science as an open enterprise [EB/OL]. [2016-03-20]. https://royalsociety.org/~/media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf.

[3] Council of European Social Science Data Archives. Depositing data – benefits [EB/OL]. [2016-03-20]. http://www.cessda.org/sharing/depositing/1/.

[4] FRY J, LOCKYER S, OPPENHEIM C, et al. Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and Research Institutes[EB/OL]. [2016-03-20]. http://repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf.

[5] United Nations Educational. Scientific and Cultural Organization. Policy guidelines for the development and promotion of open access [EB/OL]. [2016-03-20]. http://unesdoc.unesco.org/images/0021/002158/215863e. pdf.

[6] STM. Brussels declaration on STM publishing[EB/OL]. [2016-03-20]. http://www.stm-assoc.org/2007_11_01_Brussels_Declaration.pdf.

[7] VINES T H, ANDREW R L, Bock D G, et al. Mandated data archiving greatly improves access to research data[J]. The FASEB journal, 2013, 27(4): 1304-1308.

[8] WARE M, MABE M. The stm report. An overview of scientific and scholarly journal publishing [EB/OL]. [2016-04-16]. http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf.

[9] PLOS ONE. Materials and software sharing [EB/OL]. [2016-04-16]. http://www.plosone.org/static/policies.action#sharing.

[10] Australian National Data Service. Data journal guide [EB/OL]. [2016-03-20]. http://ands.org.au/guides/ data-journals.html.

[11] MAYERNIK M S, CALLAGHAN S, LEIGH R, et al. Peer review of datasets: When, why, and how [J]. Bulletin of the American Meteorological Society, 2015, 96(2): 191-201.

[12] The journal of open archaeology data (JOAD) [EB/OL]. [2016-03-20]. http://openarchaeologydata.metajnl.com/.

[13] The Journal of Open Archaeology Data. Editorial policies,

知识管理论坛，2016(3):182–191
DOI：10.13266/j.issn.2095–5472.2016.023

peer review process [EB/OL]. [2016-03-20]. http://openarchaeologydata.metajnl.com/about/editorialPolicies#peerReviewProcess.

[14] FORCE11. Joint declaration of data citation principles – FINAL [EB/OL]. [2016-03-20]. https://www.force11.org/group/joint-declaration-data-citation-principles-final

[15] European Commission. Standardization in the area of innovation and technological development, notably in the field of text and data mining [EB/OL]. [2016-03-20]. http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf.

[16] The article of the future: creating an optimal way to communicate research [EB/OL]. [2016-03-20]. http://www.articleofthefuture.com/.

[17] DataCite. Why cite data? [EB/OL]. [2016-03-20]. http://www.datacite.org/whycitedata.

Research on Publishers' Policies for Open Sharing of Scientific Data

Cui Yan[1,2]
[1]Beijing Jiaotong University Library, Beijing 100044
[2]National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract: [Purpose/significance] Publishers are key stakeholders in the process of openly sharing scientific data. Publishers cooperate with funders, data centers and research institutions to contribute to a culture of openly sharing high-quality research data. Whereas publishers have placed muchemphasis on open access to scientific data, recently they are engaged with research data in particular. **[Method/process]** This study provided an overview of the current policies, practices and provided policy recommendations for publishers. **[Result/conclusion]** These recommendations will assist publishers in furthering the goals of open access to scientific data in their organizations and networks.

Keywords: scientific data open sharing　publisher　policy